

(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION
EN MATIÈRE DE BREVETS (PCT)

(19) Organisation Mondiale de la Propriété
Intellectuelle
Bureau international



554 303

(43) Date de la publication internationale
9 décembre 2004 (09.12.2004)

PCT

(10) Numéro de publication internationale
WO 2004/107652 A1

(51) Classification internationale des brevets⁷ : H04L 12/24
(21) Numéro de la demande internationale :
PCT/FR2004/050205

(22) Date de dépôt international : 25 mai 2004 (25.05.2004)

(25) Langue de dépôt : français

(26) Langue de publication : français

(30) Données relatives à la priorité :
03 50181 27 mai 2003 (27.05.2003) FR

(71) Déposant (pour tous les États désignés sauf US) :
FRANCE TELECOM [FR/FR]; 6, place d'Alleray,
F-75015 PARIS (FR).

(72) Inventeurs; et

(75) Inventeurs/Déposants (pour US seulement) : DOUS-
SON, Christophe [FR/FR]; 88, rue des Frères Lagadec,
F-22300 LANNION (FR). CLEROT, Fabrice [FR/FR];
Le Colven, F-22700 LOUANNEC (FR). FESSANT,
Françoise [FR/FR]; 33, rue de Goas an Abat, F-22700
PERROS GUIREC (FR).

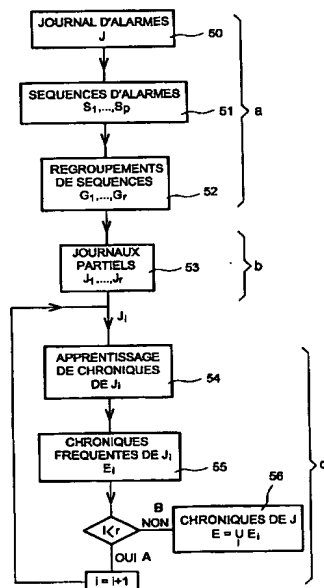
(74) Mandataire : POULIN, Gérard; BREVALEX, 3, rue du
Docteur Lancereaux, F-75008 PARIS (FR).

(81) États désignés (sauf indication contraire, pour tout titre de
protection nationale disponible) : AE, AG, AL, AM, AT,
AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO,
CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB,
GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG,

[Suite sur la page suivante]

(54) Title: METHOD FOR THE MACHINE LEARNING OF FREQUENT CHRONICLES IN AN ALARM LOG FOR THE MONITORING OF DYNAMIC SYSTEMS

(54) Titre : PROCÉDE D'APPRENTISSAGE AUTOMATIQUE DE CHRONIQUES FREQUENTES DANS UN JOURNAL D'ALARMES POUR LA SUPERVISION DE SYSTEMES DYNAMIQUES



A...YES

B...NO

50...ALARM LOG J

51...SEQUENCES OF ALARMS S₁, ..., S_p

52...REORGANISATION OF SEQUENCES G₁, ..., G_R

53...PARTIAL LOGS J₁, ..., J_R

54...LEARNING OF CHRONICLES FROM J_i

55...FREQUENT CHRONICLES FROM J_i E_i

56...CHRONICLES FROM J E = U E_i

(57) Abstract: The invention relates to a method for the machine learning of frequent chronicles in a log of alarms from a dynamic system, for the monitoring of said system, and a learning system which is used to carry out the method in a monitoring system. According to the invention, sequences of alarms (51) are selected S₁, S₂, ..., S_p from an alarm log J (50) and reorganised (52) into groups of similar sequences G₁, G₂, ..., G_R. The alarms from the groups are then used to produce (53) partial logs J₁, J₂, ..., J_R. Subsequently, chronicle learning is performed (54) on each transmitted partial log J_i, and the partial set E_i of the frequent chronicles from J_i is determined (55). Finally, a set E of chronicles from log J is formed (56) with chronicles from different partial sets E_i.

(57) Abrégé : Procédé d'apprentissage automatique de chroniques fréquentes dans un journal d'alarmes issues d'un système dynamique pour la supervision de ce système et système d'apprentissage pour la mise en oeuvre du procédé dans un système de supervision. Selon le procédé, dans un journal J d'alarmes (50), des séquences d'alarmes (51) sont sélectionnées S₁, S₂, ..., S_p puis regroupées (52) en groupes de séquences similaires G₁, G₂, ..., G_R. Les alarmes des groupes servent ensuite à fabriquer (53) les journaux partiels J₁, J₂, ..., J_R. Un apprentissage de chroniques fréquentes est alors effectué (54) sur chaque journal partiel J_i transmis et l'ensemble partiel E_i des chroniques fréquentes de J_i est déterminé (55); enfin un ensemble E de chroniques du journal J est constitué (56) avec des chroniques des divers ensembles partiels E_i.

WO 2004/107652 A1



KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

- (84) États désignés (sauf indication contraire, pour tout titre de protection régionale disponible) : ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), eurasién (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), européen (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Publiée :

- avec rapport de recherche internationale
- avant l'expiration du délai prévu pour la modification des revendications, sera republiée si des modifications sont reçues

En ce qui concerne les codes à deux lettres et autres abréviations, se référer aux "Notes explicatives relatives aux codes et abréviations" figurant au début de chaque numéro ordinaire de la Gazette du PCT.

**PROCÉDÉ D'APPRENTISSAGE AUTOMATIQUE DE CHRONIQUES
FRÉQUENTES DANS UN JOURNAL D'ALARMES POUR LA
SUPERVISION DE SYSTÈMES DYNAMIQUES**

5

DESCRIPTION

DOMAINE TECHNIQUE

La présente invention a pour objet un procédé et un système d'apprentissage automatique de chroniques fréquentes d'un journal d'alarmes d'un système dynamique pour la supervision de ce dernier.

Les systèmes dynamiques concernés par l'invention sont, par exemples, les réseaux de télécommunication, les réseaux informatiques ou toutes autres installations industrielles dont les équipements sont supervisés telles que des centrales nucléaires, des chaînes de montage, des usines automatisées etc.

La supervision d'un système dynamique consiste à surveiller son bon fonctionnement, à collecter les informations sur son état ou sur celui des composants du système, à détecter et identifier les dysfonctionnements qui peuvent se produire. Cette supervision est le plus souvent effectuée par un système informatique qui centralise des informations envoyées au cours du temps par des composants du système dynamique. Les informations reçues par le système de supervision peuvent être très diverses : par exemple des informations de déroulement de procédures, des messages d'alertes ; elles sont souvent liées à des mesures de capteurs physiques.

Le système de supervision reçoit des informations sous forme d'alarmes, chaque alarme étant formée d'un évènement d'un type donné, par exemple tel équipement électrique du système dynamique est hors
5 tension (sous forme d'un message codé), associé à sa date d'occurrence (souvent sous forme d'un nombre entier d'unités de temps). Dans le cas d'un réseau de télécommunication des exemples de types d'alarmes sont
10 « perte de signal » ou encore « perte de trame de transmission », que l'on peut encore regrouper sous le type plus général de « panne de transmission ».

Les alarmes reçues par le système de supervision sont stockées dans un journal d'alarmes qui correspond à une liste des alarmes reçues ordonnée dans
15 le temps suivant les dates d'occurrence entre une date de début et une date de fin du journal.

Un système de supervision peut recevoir un nombre considérable d'alarmes, avec de fortes variations au cours du temps : on peut par exemple
20 passer de plusieurs centaines de messages par secondes à quelques dizaines ou moins. Certaines des alarmes reçues ne sont pas indépendantes mais résultent de « cascades » d'alarmes du fait de l'interdépendance de certains composants du système dynamique supervisé.

25 L'analyse du journal d'alarmes, notamment pour rechercher les véritables causes de dysfonctionnements afin de proposer une réaction appropriée (préventive ou corrective), est une tâche difficile car il faut isoler dans la masse
30 d'informations du journal les groupes pertinents d'alarmes. Une représentation possible pour ces groupes

d'alarmes utilise des ensembles d'évènements reliés par des contraintes temporelles (sous forme de graphes) : ce sont les chroniques. La connaissance sur l'évolution d'un système dynamique peut être représentée par de
5 telles chroniques car on peut considérer que chaque chronique constitue un scénario possible pour l'évolution du système. Cette connaissance acquise via les chroniques permet donc d'anticiper le comportement du système dynamique et autorise ainsi un meilleur
10 contrôle de celui-ci.

Dans le cas d'un réseau de télécommunication par exemple, les alarmes sont générées automatiquement par les divers équipements du réseau (commutateurs, multiplexeurs, brasseurs,...) et
15 sont transmises à un superviseur central. Le flot d'alarmes contient alors des alarmes dues aux automatismes du réseau, et que l'on peut qualifier de normales, et des alarmes liées à des dysfonctionnements ; si une chronique correspond à un
20 dysfonctionnement alors ses alarmes seront analysées pour trouver l'origine de ce dysfonctionnement et y remédier. La plupart des systèmes de supervision et de contrôle ont une architecture en trois modules, 14, 16 et 15, comme illustré sur la Figure 1 : un système de
25 supervision 17 est relié à un système dynamique 10, en interaction avec l'extérieur 11, dont des composants sont munis de capteurs 12 et qui peut être commandé par des actionneurs 13 ; les capteurs 12 envoient des signaux à un module de détection 14 qui génère des
30 alarmes à partir de ces signaux et les transmet à un module de diagnostic 16 qui interprète les alarmes,

identifie les situations caractéristiques de l'évolution de 10, qui localise les composants de 10 impliqués dans ces situations, qui détermine les causes d'éventuels dysfonctionnements et qui transmet ces informations à un module de décision 15 qui détermine alors les actions à accomplir (pour viser un objectif donné ou pour ramener le système supervisé à une situation normale) sur les composants de 10 et transmet des commandes en conséquence aux actionneurs 13 du système dynamique.

L'apprentissage des chroniques intervient au niveau du module de diagnostic et permet de dégager les informations pertinentes qui sont dispersées dans le journal d'alarmes. L'identification des situations caractéristiques rencontrées au cours de l'évolution du système dynamique, notamment celles liées à des anomalies, et la détection des causes de ces situations à des fins de diagnostic sont basées sur les chroniques découvertes lors de l'apprentissage. Les chroniques peuvent aussi permettre d'anticiper certains comportements du système dynamique.

La découverte des chroniques d'un journal d'alarmes est donc une étape essentielle pour la supervision d'un système dynamique.

Si une alarme est représentée sous la forme d'un couple (A, t_A) , où A désigne un type d'évènement et t_A sa date d'occurrence, alors la contrainte temporelle « de A à B » entre deux alarmes (A, t_A) et (B, t_B) (ou encore la contrainte « de t_A à t_B »), représentée par un intervalle de temps $[t^-, t^+]$ placé entre les évènements A

et B, signifie que l'on a la relation suivante sur les dates d'occurrence :

$$t^- \leq (t_B - t_A) \leq t^+ ,$$

l'absence de contrainte entre deux instants étant
5 représenté par la contrainte $[-\infty, +\infty]$.

Une chronique (ou scénario ou encore motif temporel) du journal d'alarmes est constituée par la donnée de k éléments, k étant la taille de la chronique, c'est-à-dire de k types d'évènements du
10 journal (ou des alarmes associées) et des contraintes temporelles entre les k dates d'occurrence correspondantes. On peut dire qu'une chronique est un ensemble de types d'évènements dont les dates d'occurrence sont contraintes, et ces contraintes
15 peuvent se représenter sous la forme d'un graphe temporel.

Un graphe de contraintes temporelles est un graphe orienté dont les sommets sont les dates et dont les arcs sont libellés par les contraintes entre ces
20 dates; par exemple, pour deux dates t_1 et t_2 , l'arc de t_1 vers t_2 est libellé par la contrainte « de t_1 à t_2 ».

Il peut exister de nombreux exemples de réalisation d'une chronique C donnée dans le journal d'alarmes, on dit alors qu'il existe plusieurs
25 instances de la chronique C ; une instance d'une chronique correspond donc à une liste des alarmes de la chronique (ou des évènements associés à ces alarmes), ordonnées dans le temps, extraite du journal.

Un exemple de chronique est illustré sur la
30 Figure 2, cette chronique met en jeu des évènements de types a (en 1 ou 4), b (en 2) et c (en 3) avec des

indications d'intervalles de temps relatifs aux contraintes temporelles (par exemple 5) : un évènement de type a (en 1) se produit à un temps initial, il précède un évènement de type c (en 3) qui se produit
5 entre 2 et 5 unités de temps plus tard, alors un évènement de type b se produit entre 3 et 10 unités de temps plus tard ainsi qu'un autre évènement de type a qui se produit entre 2 et 10 unités de temps plus tard (après l'évènement initial), les évènements ultérieurs
10 de types b et a se produisant respectivement entre 1 et 6 unités de temps et entre 0 et 8 unités de temps après l'évènement de type c (en 3).

Considérons alors, à titre d'exemple et à propos de la chronique de la Figure 2, une liste
15 suivante d'évènements e1 à e8 dans laquelle, par exemple, l'évènement e1 correspondant à l'occurrence d'une alarme de type a à une date $t = 4$ unités de temps est noté $e1(a, t=4)$: $e1(a, t=4), e2(d, t=5), e3(a, t=6),$
 $e4(c, t=8), e5(b, t=10), e6(e, t=11), e7(a, t=12), e8(b, t=14),$
20 liste qui fait apparaître huit alarmes (liées aux huit évènements e1 à e8) de cinq types différents notés a, b, c, d, et e. On peut, sur cet exemple, reconnaître quatre instances de la chronique de la Figure 2, à savoir $\{e1, e4, e5, e7\}, \{e3, e4, e5, e7\}, \{e3, e4, e7, e8\}$ et
25 $\{e1, e4, e7, e8\}$ qui mettent en jeu les trois types d'alarmes a, b et c et vérifient bien les relations entre dates d'occurrence de la chronique.

On appelle fréquence d'une chronique le nombre d'instances de cette chronique dans le journal
30 d'alarmes. Il s'agit donc d'un nombre d'occurrences de la chronique en réalité plutôt qu'une véritable

fréquence ; cependant on obtient trivialement une vraie fréquence (ou taux d'apparition moyen) en divisant ce nombre d'occurrences par la durée du journal d'alarmes, c'est-à-dire la différence entre ses dates de fin et de début, car l'analyse du journal se fait à une date de fin donnée. La taille du journal d'alarmes (ou sa longueur) est le nombre d'alarmes qu'il contient. La taille d'une chronique est le nombre d'évènements dont elle est formée c'est-à-dire la taille de ses instances.

Une chronique est dite fréquente, dans un journal, lorsque sa fréquence dans le journal dépasse une fréquence seuil f_{min} donnée.

Le processus d'apprentissage de chroniques fréquentes dans un journal d'alarmes correspond à l'exploration et à l'analyse du journal pour découvrir les chroniques dont la fréquence des instances dans le journal dépasse une fréquence seuil donné. Il s'agit en effet d'explorer des séquences d'alarmes à la fois sur le plan des évènements et sur le plan des contraintes temporelles entre leurs dates d'occurrence pour trouver les chroniques, mais aussi de reconnaître des chroniques identifiées, à travers leurs instances, au sein du journal (le nombre de fois où une chronique est reconnue dans le journal étant égal à sa fréquence).

ETAT DE LA TECHNIQUE ANTERIEURE

Il existe plusieurs algorithmes [1,2,3,4] capables de construire et de découvrir toutes les chroniques fréquentes présentes dans un journal

d'alarmes, de nombreuses variantes de ces algorithmes ont été développées, notamment le système FACE [4,6] (acronyme pour « Frequency Analyser for Chronicle Extraction »).

5 L'apprentissage automatique de chroniques sur un ordinateur est cependant très coûteux en temps de calcul et en occupation mémoire, de plus, trop
diminuer la fréquence seuil (pour trouver des chroniques plus rares) peut aboutir à un phénomène
10 d'explosion combinatoire qui sature le ordinateur. Ce coût important est le reflet de la complexité du processus, cette dernière est principalement liée à deux facteurs : (i) le nombre d'alarmes présentes dans le journal (la taille du journal à traiter) et, (ii) la
15 fréquence seuil f_{min} , fixée par l'utilisateur, qui fixe la fréquence minimale des chroniques à rechercher dans le journal.

On notera qu'en ce qui concerne la gestion des contraintes temporelles, un graphe de contraintes
20 temporelles peut avoir plusieurs représentations équivalentes (il y a donc autant de représentations équivalentes d'une chronique correspondante) mais qu'il existe une seule représentation minimale (au sens d'une relation d'ordre partiel) ; le calcul de cette dernière
25 représentation, ainsi que la vérification de sa consistance globale, est généralement effectué par un algorithme bien connu de type Floyd-Warshall d'une complexité en $O(n^3)$ où n est le nombre d'instantanés (ou dates) du graphe [5] et est donc lié au nombre
30 d'alarmes.

En pratique, l'homme du métier fixe à la fois la taille maximum L_{\max} des contraintes temporelles, c'est-à-dire la durée maximale entre les dates d'occurrence des alarmes d'une chronique (ou durée de la chronique), et la fréquence seuil f_{\min} pour effectuer dans un temps raisonnable un apprentissage de chroniques sur un journal d'alarmes J donné. Ce « temps raisonnable » ne sera évidemment pas le même si la supervision et le contrôle du système dynamique se font en temps réel ou si l'analyse du journal est faite en différé, pour des besoins d'acquisition d'expertise par exemple.

Le résultat de l'apprentissage est alors l'ensemble E des chroniques du journal $E = \{C_1, \dots, C_M\}$ de fréquences supérieures ou égales à f_{\min} et de durées inférieures ou égales à L_{\max} comme indiqué sur la Figure 3.

Figure 3 sur laquelle sont indiquées les étapes de l'apprentissage typiques de l'art antérieur : un journal d'alarmes 30 est transmis à un module d'apprentissage de chroniques fréquentes 31, à fréquence seuil f_{\min} et à taille maximum des contraintes temporelles L_{\max} 30, qui produit alors les chroniques fréquentes de J : C_1, \dots, C_M en 32.

25

Il arrive souvent que l'homme du métier ne puisse pas diminuer la fréquence seuil f_{\min} (le second facteur de complexité (ii) évoqué plus haut) en dessous d'un niveau relativement élevé à cause des contraintes de temps du processus d'apprentissage, dans ce cas le risque important c'est de laisser passer des chroniques

30

'rares', de fréquences inférieures à f_{\min} , mais qui peuvent être très importantes pour la bonne supervision d'un système dynamique. De tels cas concernent particulièrement les épiphénomènes. Un épiphénomène
5 étant l'apparition d'un motif dans un journal d'alarmes à un taux élevé mais sur des intervalles de temps courts par rapport à la durée du journal, en moyenne le taux d'apparition du motif est par contre très faible et donc aussi sa fréquence, notamment si on la compare
10 à f_{\min} , un tel cas est illustré sur la Figure 4. L'art antérieur ne permet pas d'apprendre les chroniques associées aux épiphénomènes en diminuant, par exemple d'un facteur dix, la fréquence seuil sans faire croître considérablement les temps de calcul.

15 Pour tenter de résoudre ce problème de détection des chroniques de basses fréquences, l'homme du métier, jouant sur le premier facteur (i) de complexité, diminue la taille du journal d'alarmes en se basant sur le fait que, statistiquement, la division
20 par un facteur donné de la taille du journal entraîne la division de la fréquence seuil par un facteur du même ordre de grandeur (pour un temps de traitement similaire). Cependant cette technique est très difficile à mettre en œuvre et peut présenter des
25 inconvénients sérieux : si des blocs d'alarmes sont supprimées du journal sans précaution, le risque de supprimer des chroniques, y compris de fréquences élevées, est très grand, ce qui dégrade la supervision et n'est guère acceptable; quant aux épiphénomènes,
30 leurs fréquences chuteront statistiquement encore plus vite que la taille du journal, ce qui ne permettra pas

davantage de les découvrir ; si des alarmes sélectionnées sont retirées du journal c'est en général sur la base d'une analyse préalable de ce journal (le plus souvent par des experts), par exemple en éliminant
5 les alarmes de certaines chroniques, dans ce cas le temps total de traitement augmente de manière importante et, en particulier, peut prohiber une supervision en temps réel.

Ces difficultés rencontrées par l'homme du
10 métier illustrent l'importance pratique considérable du facteur temps pour l'apprentissage et il est clair que le besoin existe de pouvoir réduire le temps d'apprentissage des chroniques, même lorsqu'on ne recherche pas d'épiphénomènes, pour la supervision des
15 systèmes dynamiques ou pour l'acquisition de connaissances (d'expertise) sur ces systèmes.

EXPOSÉ DE L'INVENTION

20 L'invention vise ainsi à améliorer la vitesse de traitement d'une application d'apprentissage de chroniques d'un journal d'alarmes de façon, en particulier, à autoriser une diminution de la fréquence seuil, pour la recherche de chroniques correspondant à
25 des épiphénomènes, tout en conservant des temps de traitement acceptables.

L'invention met en œuvre des techniques d'analyse de données comme celles relevant de la classification de données (ou « clustering » en
30 anglais) c'est-à-dire des méthodes de groupement de données. Il existe de très nombreux algorithmes qui

calculent de tels agrégats ou groupes ou encore grappes de données (des « clusters » en anglais).

L'invention concerne un procédé d'apprentissage automatique comportant un prétraitement du journal d'alarmes, ne présentant pas les inconvénients de l'art antérieur, qui permet de fabriquer des journaux d'alarmes partiels (de tailles réduites), à partir du journal d'alarmes d'origine, sur chacun desquels un apprentissage est ensuite effectué.

5 Ce 'découpage' automatique en journaux partiels doit être 'intelligent', en effet un découpage systématique ou aléatoire du journal d'alarmes en blocs d'alarmes n'améliore rien du point de vue des fréquences si la répartition des motifs est aléatoire : ce genre de

10 découpage ne permet pas de faire un apprentissage en basse fréquence car la fréquence d'un épiphénomène chutera plus vite que la taille du journal.

Le prétraitement de découpage automatique selon l'invention permet de s'assurer que les journaux partiels seront riches en séquences d'alarmes qui se ressemblent, ou sont similaires, du point de vue des alarmes produites.

20

En particulier, pour ce qui concerne les épiphénomènes, ce prétraitement tend à sélectionner les zones du journal d'alarmes correspondant aux pics de

25 taux d'apparition (voir sur la Figure 4, les pics 41,42,43) tout en diminuant la taille du journal à analyser, ce qui permet d'augmenter la fréquence effective d'un épiphénomène dans le nouveau journal et

30 le rend détectable dès que cette dernière fréquence atteint ou dépasse la fréquence seuil, alors qu'avec un

apprentissage fait directement sur le journal d'origine l'épiphénomène aurait eu une fréquence trop faible: en réduisant la taille, en procédant selon l'invention, on fait passer la fréquence de la chronique rare,
5 correspondant au type de similarité du journal partiel, au dessus du seuil.

Le phénomène essentiel qui est à la base de l'invention c'est que si deux parties, ou séquences d'alarmes, du journal d'alarmes d'origine contiennent
10 des instances de mêmes chroniques alors ces parties doivent être relativement similaires au sens où elles doivent contenir plusieurs éléments en commun et dans un ordre plus ou moins semblable : si chacune de ces parties est décrite par un ensemble de paramètres
15 (constituant une représentation de la partie), associés à divers aspects de son contenu en alarmes, alors la similarité des parties se traduit par la similarité (ou proximité) des ensembles de paramètres, représentatifs de ces parties, dans l'espace des paramètres. Par
20 contre, si ces parties n'ont pas de chroniques en commun, alors le découpage en parties selon l'invention n'apportera rien de plus par rapport à un apprentissage direct : cependant, dans ce cas, l'avantage de l'invention est qu'elle ne permet pas d'apprendre par
25 erreur.

Plus précisément, dans un mode dit général, l'invention est un procédé d'apprentissage automatique de chroniques fréquentes d'un journal d'alarmes d'un système dynamique, pour la supervision de ce système,
30 les alarmes étant associées à une pluralité

d'évènements du système d'une pluralité de types, caractérisé en ce qu'il comporte les étapes suivantes :

a) de sélection et de regroupement automatiques de séquences d'alarmes du journal
5 d'alarmes de façon à former des groupes de séquences d'alarmes similaires; et

b) de génération automatique d'un journal d'alarmes partiel pour chaque groupe de séquences d'alarmes similaires obtenu à l'étape a), à partir des
10 alarmes appartenant aux séquences de ce groupe; et

c) d'apprentissage automatique de chroniques fréquentes de chaque journal d'alarmes partiel obtenu à l'étape b) de façon à générer un ensemble partiel de chroniques fréquentes pour chaque
15 journal d'alarmes partiel obtenu à l'étape b), et de fabrication d'un ensemble de chroniques fréquentes du journal d'alarmes à partir des chroniques fréquentes de chacun des ensembles partiels de chroniques fréquentes obtenus.

20

Il est possible, par exemple, de former l'ensemble de chroniques fréquentes du journal d'alarmes par union ensembliste des chroniques des ensembles partiels de chroniques fréquentes, ce qui
25 présente l'avantage de n'éliminer aucune des chroniques obtenues lors de chaque apprentissage de journal partiel.

D'autres choix sont bien sûr possibles, l'homme du métier peut, par exemple, décider d'éliminer
30 des chroniques dont la fréquence est inférieure à une fréquence donnée (notamment lorsque la fréquence seuil

n'est pas la même pour l'apprentissage des divers journaux partiels), ou encore d'éliminer des chroniques contenant certains types d'alarmes, etc.

Ce mode général est représenté sur la Figure 5 : dans un journal J d'alarmes 50, des séquences d'alarmes 51 sont sélectionnées S_1, S_2, \dots, S_p puis regroupées 52 en groupes de séquences similaires G_1, G_2, \dots, G_r , réalisant ainsi l'étape a) du mode général. Les alarmes des groupes servent ensuite à fabriquer 53 les journaux partiels J_1, J_2, \dots, J_r selon l'étape b) du mode général. Puis, conformément à l'étape c) du mode général, un apprentissage de chroniques fréquentes 54 est effectué sur chaque journal partiel J_i transmis et l'ensemble partiel E_i des chroniques fréquentes de J_i est déterminé 55 ; enfin un ensemble E de chroniques de journal J est constitué 56 à partir des chroniques des divers ensembles partiels E_i .

La comparaison avec la Figure 3 fait clairement ressortir les différences de l'apprentissage de chroniques selon l'art antérieur et selon l'invention : l'apprentissage direct du journal J, en 30 et 31, pour obtenir la totalité des chroniques fréquentes en 32, est remplacé par un séquençage de J, en 51, suivi d'un regroupement de séquences similaires, en 52, permettant de fabriquer des journaux partiels 53 sur lesquels sont effectués autant d'apprentissages 'partiels', en 54 et 55, pour ensuite déterminer les chroniques fréquentes du journal J d'origine à partir des chroniques des journaux partiels 56. Le regroupement automatique de séquences similaires est un

point essentiel de l'invention car c'est lui qui permet d'obtenir des journaux partiels potentiellement riches en chroniques fréquentes.

5 L'invention concerne aussi un dispositif pour la mise en œuvre du nouveau procédé décrit plus haut, c'est-à-dire un système d'apprentissage automatique de chroniques fréquentes d'un journal d'alarmes d'un système dynamique, pour la supervision
10 de ce système, comportant des moyens d'acquisition d'alarmes du système dynamique et de génération d'un journal d'alarmes à partir des alarmes acquises, chaque alarme étant associée à un évènement du système dynamique parmi une pluralité d'évènements d'une
15 pluralité de types et à une date d'occurrence, des moyens de transmission du journal d'alarmes ainsi que des moyens d'apprentissage de chroniques aptes à mettre en œuvre une méthode d'apprentissage automatique de chroniques d'un journal d'alarmes, de fréquences
20 supérieures ou égales à un seuil de fréquence minimum f_0 réglable et de durée maximum T réglable, et aptes à transmettre les chroniques obtenues, caractérisé en ce qu'il comporte en outre :

- un module de sélection et de regroupement
25 de séquences d'alarmes apte à recevoir un journal d'alarmes et apte à sélectionner et à regrouper des séquences d'alarmes du journal d'alarmes, et apte à former un groupe de séquences d'alarmes similaires et à transmettre ce groupe;

30 - un module de fabrication d'un journal d'alarmes partiel à partir des alarmes d'un groupe de

séquences d'alarmes similaires reçu du module de sélection et de regroupement de séquences d'alarmes du journal d'alarmes, le module étant apte à transmettre le journal d'alarmes partiel obtenu aux moyens d'apprentissage de chroniques;

- un module de fabrication d'un ensemble de chroniques fréquentes du journal d'alarmes, à partir des chroniques transmises par les moyens d'apprentissage de chroniques, le module étant apte à transmettre les chroniques de l'ensemble de chroniques fréquentes.

Ce système d'apprentissage selon l'invention est illustré sur la Figure 6 : le système d'apprentissage 66 est ici représenté en liaison avec le module de diagnostic 16 d'un système de supervision 17 d'un système dynamique 10. Des alarmes 60 sont transmises à des moyens 61 d'acquisition d'alarmes et de génération d'un journal d'alarmes J. Le journal J est transmis à un module M2 de sélection et de regroupement de séquences d'alarmes 62. Le module sélectionne des séquences S_1, S_2, \dots, S_p (en nombre p variable) et forme des groupes de séquences similaires G_1, G_2, \dots, G_r (en nombre r variable); il transmet chaque groupe G_k , pour k variant de 1 à r, à un module M3 de fabrication d'un journal d'alarmes partiel 63. Chaque journal partiel J_k , pour $1 \leq k \leq r$, est généré à partir des alarmes des séquences du groupe G_k correspondant. Le module M3 transmet chaque journal partiel J_k aux moyens M4 d'apprentissage de chroniques fréquentes 64, de fréquence minimum f_0 et de durée

maximum T réglables, qui produisent alors un ensemble partiel E_k de chroniques fréquentes de J_k contenant un nombre $M(k)$, variable suivant l'indice k , de chroniques désignées par $C_{1(k)}, \dots, C_{M(k)}$. Enfin, un module M5 de fabrication d'un ensemble de chroniques 65, génère un ensemble E de chroniques de J à partir des chroniques de tous les ensembles partiels E_k transmis par le module M4. Le module M5 transmet alors les chroniques C_1, \dots, C_M de E pour leur exploitation par le système de supervision. L'homme du métier notera que, dans les systèmes d'apprentissage des systèmes de supervision de l'art antérieur, les modules M2, M3 sont absents et que le module M1 transmet directement le journal J au moyens d'apprentissage M4 qui fabriquent directement l'ensemble E (le module M5 est donc aussi absent).

BREVE DESCRIPTION DES DESSINS

Les caractéristiques et avantages de l'invention exposée ci-dessus, ainsi que d'autres qui ressortiront de la description suivante de modes particuliers de réalisation, donnés à titre d'exemples, apparaissent davantage en se référant aux dessins annexés, sur lesquels :

- la Figure 1 est un schéma logique d'un système de supervision d'un système dynamique comme indiqué ci-dessus.

- la Figure 2 représente un exemple de chronique mettant en jeu des événements de trois types a, b et c.

- la Figure 3 est un schéma-bloc montrant l'apprentissage de chroniques fréquentes de l'art antérieur.

5 - la Figure 4 représente une courbe du taux d'apparition σ (en ordonnée) d'un motif, en fonction du temps t (en abscisse), typique d'un épiphénomène. Des pics caractéristiques 41, 42 et 43 sont indiqués.

10 - la Figure 5 est un schéma-bloc montrant l'apprentissage de chroniques fréquentes selon l'invention dans son mode général.

- la Figure 6 est un schéma-bloc montrant un système d'apprentissage selon l'invention, dans un système de supervision d'un système dynamique.

15

EXPOSÉ DÉTAILLÉ DE MODES DE RÉALISATION PARTICULIERS

La sélection automatique de séquences d'alarmes au sein du journal d'alarmes, à l'étape a) du mode général de l'invention, est une étape importante du procédé qui peut être effectuée de bien des façons :
20 par exemple, il est possible de prendre autant de séquences que d'alarmes dans le journal, chaque séquence contenant donc une alarme, mais dans ce cas le nombre de séquences à regrouper est alors considérable.

25 Avantageusement, dans un mode dit avec découpage, à l'étape a) du procédé d'apprentissage de chroniques selon l'invention, la sélection automatique de séquences d'alarmes est effectuée au moyen d'un découpage automatique du journal d'alarmes en parties,
30 chaque partie étant formée d'alarmes du journal d'alarmes dont les dates d'occurrence sont ordonnées

dans le temps et sont comprises entre une date de début et une date de fin associées à cette partie du journal, chaque partie du journal d'alarmes définissant une séquence d'alarmes sélectionnée dont les alarmes sont
5 celles qui appartiennent à cette partie. Ce mode peut donc être mis en œuvre pour réaliser la sélection de séquences de l'étape a) du mode général.

Le découpage du journal d'alarmes en
10 parties, ou par tranches de temps, peut être réalisé selon de nombreuses modalités différentes : à titre d'exemples non limitatifs, les parties peuvent être de tailles (en nombre d'alarmes contenues) différentes ou bien identiques, elles peuvent avoir la même durée ou
15 bien des durées variables, elles peuvent être disjointes ou bien avoir des alarmes en commun, elles peuvent constituer dans leur ensemble une partition du journal d'alarmes ou bien, au contraire, ne pas prendre en compte certaines alarmes, etc.

20 Avantageusement, l'union ensembliste des parties du journal d'alarmes reconstituera le journal de façon à ne pas perdre des alarmes pouvant appartenir à des instances de chroniques à ce stade du procédé d'apprentissage : toute alarme du journal appartiendra
25 alors à au moins une des parties du journal.

Selon un mode particulier de l'invention, dit mode avec découpage complet, le découpage précédent du journal d'alarmes en parties est tel que toute alarme du journal d'alarmes appartient à au moins une
30 des parties du journal d'alarmes. Ce mode présentant l'avantage de n'éliminer aucune alarme du journal

d'alarmes lors du découpage. Ce mode, mis en œuvre avec le mode avec découpage, peut être utilisé pour réaliser la sélection de séquences à l'étape a) du mode général.

5 Le regroupement, à l'étape a), des séquences sélectionnées qui présentent des similarités peut être effectué à l'aide d'une mesure de similarité sur un espace dans lequel chaque séquence d'alarmes est décrite par un ensemble de paramètres qui peut être vu
10 comme définissant des coordonnées d'un point dans cet espace appelé espace de représentation des séquences. Chaque paramètre d'une séquence, ou coordonnée du point représentatif de la séquence dans l'espace des représentations, est associé à la description du
15 contenu en alarmes de la séquence en termes d'un type d'alarme donné. Ainsi, si l'on dispose de A types distincts d'alarmes pour décrire le contenu des séquences d'alarmes, l'espace de représentation des séquences est à A dimensions. La valeur d'une
20 coordonnée d'une séquence correspond quant à elle au poids de la séquence en alarmes du type associé à cette coordonnée.

 L'importance du choix d'une représentation des données, préalablement aux opérations de
25 regroupement de ces données, est bien connue des spécialistes en classification de données. Il existe un très grand nombre de représentations possibles pour des données, suivant les caractéristiques qui leur sont attachées : cette variété de représentations se traduit
30 en particulier par des dimensions différentes de l'espace de représentation.

L'un des modes particuliers de l'invention utilise une représentation, dite pondérée, dans laquelle chaque séquence d'alarmes sélectionnée à l'étape a) du procédé est représentée, dans l'espace de
5 représentation de dimension A, par un point ayant A coordonnées, la coordonnée de rang j, où j désigne un indice entier quelconque compris entre 1 et A, est égale au nombre de fois où le type d'alarme associé à l'indice j figure dans la séquence d'alarmes.

10

L'homme du métier sait que la mesure de similarité, mentionnée plus haut, utilisée pour calculer les regroupement de séquences à partir des mesures de 'proximité' des séquences dans l'espace de
15 représentation, notamment dans les algorithmes de groupement des méthodes de classification, est à prendre au sens mathématique c'est-à-dire que cette mesure n'est pas obligatoirement une distance (elle peut ne pas respecter l'inégalité triangulaire et on
20 parle alors de semi-métrique, par exemple la mesure en cosinus qui mesure le cosinus de l'angle entre deux vecteurs dont les composantes sont les coordonnées, respectivement, des points représentatifs de deux séquences considérées). Cependant, le plus souvent,
25 cette mesure de similarité est une distance et permet donc de munir l'espace de représentation d'une métrique. A titre d'exemples, non limitatifs, de métriques citons la métrique (ou distance) de Minkowski (qui contient comme cas particuliers la distance
30 Euclidienne et la distance 'City-block' encore appelée

distance de Manhattan), la distance de Mahalanobis, la distance de Chebychev, etc...

Les points, représentant les séquences d'alarmes, les plus proches entre eux dans l'espace de représentation des séquences forment des grappes et
5 chaque grappe correspond à un groupe de séquences dites similaires. Le critère de proximité de deux points de l'espace de représentation étant que si la mesure de similarité de ces points est inférieure ou égale à une
10 valeur seuil donnée alors les deux points appartiennent à la même grappe, et si la mesure de similarité de ces points est supérieure à la valeur seuil alors les deux points appartiennent à des grappes distinctes.

Il existe à ce jour de très nombreux
15 algorithmes de groupement de données, certains assurent même l'invariance (par transformations linéaires) des grappes grâce à une normalisation préalable des données.

Dans un mode de réalisation avantageux de
20 l'invention, dit mode avec groupement de séquences, à l'étape a) du procédé, le regroupement automatique des séquences d'alarmes, pour former des groupes de séquences d'alarmes similaires, est réalisée au moyen d'une méthode de groupement. Ce mode peut donc être mis
25 en œuvre pour réaliser le regroupement à l'étape a) du mode général, et il peut aussi être utilisé conjointement soit avec le mode avec découpage soit avec le mode avec découpage complet pour la réalisation de l'étape a) du mode général.

30 Le choix de la représentation des séquences d'alarmes conditionne à la fois la pertinence des

regroupements de séquences et la complexité des calculs de mesures de similarité à effectuer (du fait du nombre de dimensions de l'espace de représentation associé). Le contenu d'une séquence d'alarmes du journal d'alarmes peut être décrit de manière plus ou moins exhaustive : par exemple si le journal contient N types d'alarmes distincts alors il est possible de décrire le contenu en alarmes de la séquence sur la base de ces N types d'alarmes, mais on peut aussi choisir de décrire ce contenu sur la base d'un nombre inférieur de types d'alarmes et dans ce cas certaines alarmes (bien que figurant toujours dans la séquence), correspondant à des types absents de la représentation, ne seront pas décrites ; dans ce dernier cas cependant, la dimension de l'espace de représentation est réduite, ce qui est avantageux du point de vue de la complexité des calculs de regroupement.

Ainsi, dans un mode particulier du procédé d'apprentissage selon l'invention, dit mode avec sélection de types, à l'étape a), la formation des groupes de séquences d'alarmes similaires est effectuée au moyen des étapes suivantes consistant à:

- représenter chacune des séquences d'alarmes du journal d'alarmes par son contenu, sur la base d'un ensemble de types d'alarmes à A éléments pris parmi les types d'alarmes distincts du journal d'alarmes, en nombre supérieur ou égal à A, dans un espace de représentation des séquences d'alarmes de dimension A; et
- regrouper automatiquement des séquences d'alarmes du journal d'alarmes, dans l'espace de

représentation des séquences d'alarmes, de façon à former des groupes de séquences d'alarmes similaires. Ce mode est dépendant du mode général ou du mode avec découpage ou du mode avec découpage complet ou du mode
5 avec groupement de séquences.

En particulier, ce mode avec sélection de types peut avantageusement utiliser la représentation pondérée des séquences d'alarmes décrite plus haut. Ce mode avec sélection de types peut donc être mis en
10 œuvre pour réaliser la formation de groupes à l'étape a) du mode général, et il peut aussi être utilisé conjointement soit avec le mode avec découpage soit avec le mode avec découpage complet soit encore avec le mode avec groupement de séquences, éventuellement
15 combiné avec l'un des deux modes précédents, pour la réalisation de l'étape a) du mode général.

La simple élimination de certains des types d'alarmes présents dans le journal d'alarmes a pour
20 inconvénient que certaines alarmes (des types éliminés) ne sont pas décrites. Pour décrire toutes les alarmes tout en réduisant le nombre de dimensions de l'espace de représentation des séquences d'alarmes, il est par exemple possible de d'abord regrouper des types
25 d'alarmes du journal d'alarmes qui sont similaires, en particulier à l'aide d'une méthode de groupement, pour fabriquer des groupes de types d'alarmes (le nombre S de groupes étant inférieur ou égal au nombre N initial de types d'alarmes), puis de décrire le contenu des
30 séquences d'alarmes sur la base de ces S groupes de types d'alarmes. Ainsi le nombre de dimension de

l'espace de représentation est réduit à S, ce qui est
avantageux pour les calculs de regroupement, et, de
plus, toute alarme d'une séquence quelconque pourra
toujours être décrite car son type figure dans au moins
5 un des S groupes de types.

Un exemple simple de réduction du nombre de
types a été donné plus haut pour un réseau de
télécommunications : les deux types d'alarmes « perte
de signal » et « perte de trame de transmission »
10 peuvent être décrits par un type de niveau supérieur
(résultant, le plus souvent, d'un lien fonctionnel
entre les types considérés) qui est « panne de
transmission ». Comme dans le cas général du traitement
de données (ici les données sont des types d'alarmes du
15 journal d'alarmes), on associe à la description des
types un espace de représentation.

L'homme du métier notera que tout ce qui
est présenté sur les séquences d'alarmes et le
découpage d'un journal d'alarmes ainsi que les divers
20 traitements de ces entités peut se transposer, par
analogie, aux types d'alarmes dès que l'on dispose
d'une représentation de ces types ou des séquences de
types.

25 Il est aussi possible d'éliminer certains
des S groupes précédents pour n'en retenir qu'un nombre
réduit S', l'espace de représentation étant alors de
dimension S'. Ces méthodes de réduction de l'espace de
représentation par regroupement de types sont alors
30 combinées avec une regroupement ultérieur des séquences
d'alarmes pour former des groupes de séquences

d'alarmes similaires permettant de générer les journaux partiels de l'étape b) du procédé d'apprentissage selon l'invention.

C'est ainsi que dans un mode particulier de réalisation du procédé d'apprentissage selon l'invention, dit mode avec regroupement de types, à l'étape a), la formation des groupes de séquences d'alarmes similaires est effectuée au moyen des étapes suivantes consistant à:

10 - regrouper automatiquement des types d'alarmes du journal d'alarmes de façon à former des groupes de types d'alarmes similaires, le résultat du regroupement étant un nombre S de groupes de types d'alarmes; et

15 - représenter chacune des séquences d'alarmes du journal d'alarmes par son contenu, sur la base de groupes de types d'alarmes, en nombre S' inférieur ou égal à S , obtenus à l'étape précédente, dans un espace de représentation des séquences d'alarmes de dimension S' ; et

20 - regrouper automatiquement de s séquences d'alarmes du journal d'alarmes, dans l'espace de représentation des séquences d'alarmes, de façon à former des groupes de séquences d'alarmes similaires.

25 Le mode avec regroupement de types peut donc être mis en œuvre pour réaliser la formation de groupes à l'étape a) du mode général, et il peut aussi être utilisé conjointement soit avec le mode avec découpage soit avec le mode avec découpage complet soit

30 encore avec le mode avec groupement de séquences, éventuellement combiné avec l'un des deux modes

précédents, pour la réalisation de l'étape a) du mode général.

En particulier, le regroupement automatique des séquences d'alarmes du mode avec regroupement de types peut consister à former automatiquement des groupes de séquences similaires du journal d'alarmes, chaque groupe de séquences similaires étant associé à un groupe de types d'alarmes et résultant de la sélection des séquences d'alarmes du journal d'alarmes dont la teneur en alarmes de mêmes types que ceux du groupe de types considéré dépasse un seuil donné pour ce groupe. Cette méthode de regroupement présente l'avantage d'être très simple à mettre en œuvre.

Dans un mode particulier de réalisation du procédé d'apprentissage selon l'invention, dit mode avec regroupement de types avec seuils, à l'étape a), la formation des groupes de séquences d'alarmes similaires est effectuée au moyen des étapes suivantes consistant à:

- regrouper automatiquement des types d'alarmes du journal d'alarmes de façon à former des groupes de types d'alarmes similaires, le résultat du regroupement étant un nombre S de groupes de types d'alarmes; et
- représenter chacune des séquences d'alarmes du journal d'alarmes par son contenu, sur la base de groupes de types d'alarmes, en nombre S' inférieur ou égal à S , obtenus à l'étape précédente, dans un espace de représentation des séquences d'alarmes de dimension S' ; et

- former automatiquement des groupes de séquences similaires du journal d'alarmes, chaque groupe de séquences similaires étant associé à un groupe de types d'alarmes et résultant de la sélection
5 des séquences d'alarmes du journal d'alarmes dont la teneur en alarmes de mêmes types que ceux du groupe de types considéré dépasse un seuil donné pour ce groupe.

Le mode de regroupement de types avec seuils peut donc être mis en œuvre pour réaliser la
10 formation de groupes à l'étape a) du mode général, et il peut aussi être utilisé conjointement soit avec le mode avec découpage soit avec le mode avec découpage complet pour la réalisation de l'étape a) du mode général.

15

Il est aussi possible, dans un mode particulier dit mode avec groupement de types, de procéder au regroupement de types d'alarmes, dans le mode avec regroupement de types ou dans le mode avec
20 regroupement de types avec seuils, en mettant en œuvre une méthode de groupement portant sur des types d'alarmes du journal d'alarmes pour former des groupes de types d'alarmes similaires. Pour cela il suffit à l'homme du métier de choisir un espace de
25 représentation de ces types d'alarmes, dont il peut d'ailleurs réduire le nombre de dimensions selon diverses méthodes décrites plus haut mais appliquées aux types, et une mesure de similarité pour pouvoir ensuite utiliser l'un des nombreux algorithmes de
30 groupement existant (citons, à titre d'exemples non limitatifs, les algorithmes de : Groupement

Hiérarchique Ascendant, « K-Means » ou K-moyennes, « Fuzzy K-Means » ou K-moyennes floues, Mélange de Gaussiennes, GTM « Générative Topographic Mapping », GSOM « Generative Self Organizing Map » [7], Carte de
5 Kohonen [9], Cartes auto-organisatrices de Kohonen [8,9], etc.) .

Dans des modes particuliers du procédé d'apprentissage selon l'invention, ce regroupement automatique des types d'alarmes au moyen d'une méthode
10 de groupement évoqué précédemment (dans le mode avec groupement de types) est avantageusement effectué soit par une méthode de groupement à base de carte sémantique des types d'alarmes (c'est-à-dire que chaque alarme est traitée comme une suite de symboles, ou
15 texte, et on effectue le groupement de ces suites) soit par une méthode de groupement basée sur le profil d'accumulation dans le temps (normalisé ou non), dans le journal d'alarmes, de chaque type d'alarmes.

20 Dans tous les modes de réalisation de l'invention, il est bien sûr possible d'effectuer, à l'étape c) du procédé, l'apprentissage de chroniques sur les journaux d'alarmes partiels obtenus à l'issue de l'étape b) du procédé par une suite d'opérations
25 exécutées en série sur un ordinateur.

Cependant, dans une réalisation avantageuse de l'invention permettant de diminuer encore le temps de traitement et compatible avec les divers modes décrits, ces opérations d'apprentissage de l'étape c)
30 sont exécutées en parallèle sur un ordinateur. Ce sont, en effet, ces opérations d'apprentissage de

chroniques qui prennent le plus de temps parmi toutes les opérations liées aux autres étapes du procédé selon l'invention ; cependant, il est clair que ces dernières opérations peuvent aussi être effectuées en parallèle
5 pour gagner du temps.

Le mode préféré de réalisation de l'invention, développé et utilisé pour la supervision d'un système de télécommunications, met en œuvre sur un
10 calculateur, pour l'étape a) du mode général, un découpage du journal d'alarmes selon le mode avec découpage dans lequel les parties ou 'tranches horaires' découpées ont toutes la même durée $\sigma = 2 \cdot T$, où T désigne la durée maximum des chroniques à
15 apprendre qui est fixée pour l'exécution de l'étape c) du procédé, le découpage étant tel que deux parties consécutives quelconques présentent un recouvrement temporel de valeur T. Ce découpage particulier est tel que si, statistiquement, une instance quelconque d'une
20 chronique (de longueur maximum T) est présente dans J alors elle est présente dans au moins une partie de J.

Le journal d'alarmes J sur lequel on applique le procédé selon l'invention est découpé en p parties couvrant la totalité des alarmes du journal, que l'on peut supposer consécutives sans restreindre la
25 généralité ; notons σ_i , pour $1 \leq i \leq p$, la i-ème partie du journal, alors les dates de début et de fin associées à cette partie, respectivement D_i et F_i , vérifient :

30
$$D_i = D + T \cdot (i-1)$$

$$F_i = D_i + \sigma,$$

où D désigne la date de début du journal d'alarmes J,

et pour la (i+1)ième partie suivante, de dates de début et de fin, respectivement, D_{i+1} et F_{i+1}
5 vérifient :

$$D_{i+1} = D + T \cdot i = D_i + T$$

$$F_{i+1} = D_{i+1} + \sigma = F_i + T .$$

Ce découpage est donc aussi conforme au
10 mode avec découpage complet comme on peut le vérifier facilement.

On procède alors à une classification de l'ensemble des σ_i , i variant de 1 à p, en r groupes G_k , $1 \leq k \leq r$, à l'aide d'une méthode de groupement à
15 base de cartes de Kohonen ; chaque groupe G_k (avec: $1 \leq k \leq r$) de parties (ou séquences d'alarmes correspondantes) similaires comprenant un nombre variable de parties (ou séquences d'alarmes). Cette dernière étape de classification relève donc du mode
20 avec groupement de séquences. Dans une variante avantageuse du mode préféré, la représentation pondérée est utilisée pour décrire les parties de J avec un espace de représentation comportant autant de dimensions que le journal J comporte de types d'alarmes
25 distincts.

A l'étape b), un journal d'alarmes partiel J_k est fabriqué pour chaque groupe correspondant G_k à partir de l'union ensembliste des parties, ou séquences d'alarmes correspondantes, du groupe. Ainsi aucune
30 séquence d'alarmes d'un groupe quelconque n'est omise.

Enfin à l'étape c), un apprentissage automatique de chroniques fréquentes est effectué au moyen du logiciel d'apprentissage FACE sur chacun des journaux d'alarmes partiels J_k , $1 \leq k \leq r$, obtenus
5 précédemment. Pour chaque journal partiel J_k on obtient donc un ensemble partiel E_k constitué des $M(k)$ chroniques fréquentes distinctes de J_k : $E_k = \{C_{1(k)}, \dots, C_{M(k)}\}$ (où $C_{m(k)}$, $1 \leq m(k) \leq M(k)$, désigne une chronique de J_k). Le résultat final étant l'ensemble E des
10 chroniques fréquentes de J , E étant formé par l'union ensembliste de tous les ensembles partiels E_k , pour $k = 1, \dots, r$.

Le logiciel FACE [4,6] est un outil d'apprentissage particulièrement bien adapté aux
15 alarmes et chroniques produites par des systèmes de télécommunication; dans des variantes du mode préféré, cependant, d'autres logiciels d'apprentissage sont mis en œuvre à l'étape c).

Les chroniques obtenues serviront ensuite
20 au module de diagnostic d'un système de supervision pour l'identification des situations caractéristiques du comportement du système dynamique supervisé.

De manière générale, dans le procédé d'apprentissage de chroniques fréquentes d'un journal
25 d'alarmes selon le mode préféré :

- on fixe une durée maximum T des chroniques à apprendre à l'étape c); et,
- à l'étape a), la différence entre la date de fin et la date de début de toute partie du journal
30 d'alarmes est égale à $2*T$; et,

les parties sont découpées dans le journal d'alarmes de façon à ce que, pour une partie quelconque donnée de date de début D' , la partie dont la date de début ultérieure D'' est la plus proche de D' , si elle
5 existe, est telle que sa date de début D'' est égale à la date D' augmentée de T ; et,
le regroupement automatique de l'ensemble des séquences d'alarmes du journal d'alarmes obtenues est réalisé au moyen d'un algorithme à base de cartes auto-
10 organisatrices de Kohonen ; et
- à l'étape b), pour chaque groupe de séquences d'alarmes similaires obtenu à l'étape a), un journal d'alarmes partiel est fabriqué à partir de l'union ensembliste des séquences d'alarmes du groupe
15 de séquences d'alarmes similaires ; et
- à l'étape c), l'apprentissage automatique des chroniques est réalisé au moyen du système d'apprentissage FACE. Le mode préféré est un mode dépendant du mode avec découpage complet et du mode
20 avec groupement de séquences.

RÉFÉRENCES

- 5 [1] : H.Mannila, H.Toivonen, A.I.Veramo : 'Discovering frequent episodes in sequences', in Proc.1st International Conference on Knowledge Discovery and Data Mining, pages 210-215, Montréal, Québec, Canada (1995).
- 10 [2] : H.Mannila, H.Toivonen : 'Multiple uses of frequent sets and condensed representations', in Proc.2nd International Conference on Knowledge Discovery and Data Mining, pages 189-194, Portland, Oregon, USA (1996).
- 15 [3] : R.Agrawal, H.Mannila, R.Srikant, H.Toivonen, A.I.Veramo : 'Fast discovery of association rules' in Advances in Knowledge Discovery and Data Mining, pages 307-328, AAAI Press/ the MIT Press (1996).
- 20 [4] : T.Vu Duong : 'Découverte de chroniques à partir de journaux d'alarmes. Application à la supervision de réseaux de télécommunications', Thèse en informatique et télécommunications, Institut National Polytechnique de Toulouse, soutenue publiquement le 28 mars 2001.
- 25 [5] : R.Dechter, I.Meiri, J.Pearl : 'Temporal constraint networks', in Artificial Intelligence, Special Volume on Knowledge Representation, **49**(1-3), 61-95 (1991). Elsevier.
- 30 [6] : C.Dousson, T.Vu Duong : 'Discovering chronicles with numerical time constraints from alarm logs for monitoring dynamic systems', in Proc. Of the 6th International Joint Conference on Artificial Intelligence (IJCAI 99), pp.620-626, (1999).
- 35 [7] : J.J.Verbeek, N.Vlassis, and B.J.A.Kröse : 'The Generative Self-Organizing Map', IAS technical report IAS-UVA-02-03, May 2002 (website :<http://www.science.uva.nl/research/ias/publications/reports/>).
- 40 [8] : M.Dittenbach, D.Merkl, A.Rauber : 'The growing hierarchical self-organizing map', in International Joint Conference on Neural Networks

(IJCNN 2000), volume vi, pages 15-19, Como, Italy, IEEE Computer Society (2000).

- 5 [9]: T.Kohonen: 'Self-Organizing Map', third edition (2002), Springer-Verlag.

REVENDICATIONS

1. Procédé d'apprentissage automatique de chroniques fréquentes d'un journal d'alarmes, les
5 alarmes étant associées à une pluralité d'évènements d'une pluralité de types, caractérisé en ce qu'il comporte les étapes suivantes :

a) de sélection et de regroupement automatiques de séquences d'alarmes du journal
10 d'alarmes de façon à former des groupes de séquences d'alarmes similaires ; et

b) de génération automatique d'un journal d'alarmes partiel pour chaque groupe de séquences d'alarmes similaires obtenu à l'étape a), à partir des
15 alarmes appartenant aux séquences de ce groupe ; et

c) d'apprentissage automatique de chroniques fréquentes de chaque journal d'alarmes partiel obtenu à l'étape b) de façon à générer un ensemble partiel de chroniques fréquentes pour chaque
20 journal d'alarmes partiel obtenu à l'étape b), et de fabrication d'un ensemble de chroniques fréquentes du journal d'alarmes à partir des chroniques fréquentes de chacun des ensembles partiels de chroniques fréquentes obtenus.

25

2. Procédé selon la revendication 1 dans lequel, à l'étape a), la sélection automatique de séquences d'alarmes est effectuée au moyen d'un découpage automatique du journal d'alarmes en parties,
30 chaque partie étant formée d'alarmes du journal d'alarmes dont les dates d'occurrence sont ordonnées

dans le temps et sont comprises entre une date de début et une date de fin associées à cette partie du journal, chaque partie du journal d'alarmes définissant une séquence d'alarmes sélectionnée dont les alarmes sont
5 celles qui appartiennent à cette partie.

3. Procédé selon la revendication 2 dans lequel le découpage du journal d'alarmes en parties est tel que toute alarme du journal d'alarmes appartient à
10 au moins une des parties du journal d'alarmes.

4. Procédé selon l'une quelconque des revendications 1 à 3 dans lequel, à l'étape a), le regroupement automatique des séquences d'alarmes, pour
15 former des groupes de séquences d'alarmes similaires, est réalisée au moyen d'une méthode de groupement.

5. Procédé selon l'une quelconque des revendications 1 à 4 dans lequel, à l'étape a), la
20 formation des groupes de séquences d'alarmes similaires est effectuée au moyen des étapes suivantes consistant à :

- représenter chacune des séquences d'alarmes du journal d'alarmes par son contenu, sur la
25 base d'un ensemble de types d'alarmes à A éléments pris parmi les types d'alarmes distincts du journal d'alarmes, en nombre supérieur ou égal à A, dans un espace de représentation des séquences d'alarmes de dimension A; et

30 - regrouper automatiquement des séquences d'alarmes du journal d'alarmes, dans l'espace de

représentation des séquences d'alarmes, de façon à former des groupes de séquences d'alarmes similaires.

6. Procédé selon la revendication 5 dans lequel chaque séquence d'alarmes sélectionnée à l'étape a) est représentée, dans l'espace de représentation de dimension A, par un point ayant A coordonnées, la coordonnée de rang j, où j désigne un indice entier quelconque compris entre 1 et A, est égale au nombre de fois où le type d'alarme associé à l'indice j figure dans la séquence d'alarmes.

7. Procédé selon l'une quelconque des revendications 1 à 4 dans lequel, à l'étape a), la formation des groupes de séquences d'alarmes similaires est effectuée au moyen des étapes suivantes consistant à :

- regrouper automatiquement des types d'alarmes du journal d'alarmes de façon à former des groupes de types d'alarmes similaires, le résultat du regroupement étant un nombre S de groupes de types d'alarmes; et

- représenter chacune des séquences d'alarmes du journal d'alarmes par son contenu, sur la base de groupes de types d'alarmes, en nombre S' inférieur ou égal à S, obtenus à l'étape précédente, dans un espace de représentation des séquences d'alarmes de dimension S'; et

- regrouper automatiquement des séquences d'alarmes du journal d'alarmes, dans l'espace de

représentation des séquences d'alarmes, de façon à former des groupes de séquences d'alarmes similaires.

8. Procédé selon l'une quelconque des
5 revendications 1 à 3 dans lequel, à l'étape a), la formation des groupes de séquences d'alarmes similaires est effectuée au moyen des étapes suivantes consistant à:

- regrouper automatiquement des types
10 d'alarmes du journal d'alarmes de façon à former des groupes de types d'alarmes similaires, le résultat du regroupement étant un nombre S de groupes de types d'alarmes; et

- représenter chacune des séquences
15 d'alarmes du journal d'alarmes par son contenu, sur la base de groupes de types d'alarmes, en nombre S' inférieur ou égal à S , obtenus à l'étape précédente, dans un espace de représentation des séquences d'alarmes de dimension S' ; et

20 - former automatiquement des groupes de séquences similaires du journal d'alarmes, chaque groupe de séquences similaires étant associé à un groupe de types d'alarmes et résultant de la sélection des séquences d'alarmes du journal d'alarmes dont la
25 teneur en alarmes de mêmes types que ceux du groupe de types considéré dépasse un seuil donné pour ce groupe.

9. Procédé selon l'une quelconque des
revendications 7 ou 8 dans lequel, à l'étape a),
30 l'étape de regroupement automatique de types d'alarmes,

pour former des groupes de types d'alarmes similaires, est réalisée au moyen d'une méthode de groupement.

10. Procédé selon la revendication 9 dans lequel, à l'étape a), le regroupement automatique des types d'alarmes similaires du journal d'alarmes est effectué au moyen d'une méthode de groupement à base de carte sémantique des types d'alarmes.

11. Procédé selon la revendication 9 dans lequel, à l'étape a), le regroupement automatique des types d'alarmes similaires du journal d'alarmes est effectué au moyen d'une méthode de groupement basée sur le profil d'accumulation dans le temps, dans le journal d'alarmes, de chaque type d'alarme.

12. Procédé selon l'une quelconque des revendications 1 à 11 dans lequel, à l'étape c), l'apprentissage sur les journaux d'alarmes partiels obtenus à l'étape b) est effectué en série.

13. Procédé selon l'une quelconque des revendications 1 à 11 dans lequel, à l'étape c), l'apprentissage sur les journaux d'alarmes partiels obtenus à l'étape b) est effectué en parallèle.

14. Procédé selon les revendications 3 et 4 dans lequel :

- on fixe une durée maximum T des chroniques à apprendre à l'étape c); et,

- à l'étape a), la différence entre la date de fin et la date de début de toute partie du journal d'alarmes est égale à $2 \cdot T$; et,
les parties sont découpées dans le journal d'alarmes de façon à ce que, pour une partie quelconque donnée de date de début D' , la partie dont la date de début ultérieure D'' est la plus proche de D' , si elle existe, est telle que sa date de début D'' est égale à la date D' augmentée de T ; et,
- 10 la regroupement automatique de l'ensemble des séquences d'alarmes du journal d'alarmes obtenues est réalisé au moyen d'un algorithme à base de cartes auto-organisatrices de Kohonen ; et,
- à l'étape b), pour chaque groupe de séquences d'alarmes similaires obtenu à l'étape a), un journal d'alarmes partiel est fabriqué à partir de l'union ensembliste des séquences d'alarmes du groupe de séquences d'alarmes similaires ; et,
- 15
- à l'étape c), l'apprentissage automatique des chroniques est réalisé au moyen du système d'apprentissage FACE.
- 20

15. Système d'apprentissage automatique de chroniques fréquentes d'un journal d'alarmes, comportant des moyens d'acquisition d'alarmes et de génération d'un journal d'alarmes à partir des alarmes acquises, chaque alarme étant associée à un évènement parmi une pluralité d'évènements d'une pluralité de types et à une date d'occurrence, des moyens de transmission du journal d'alarmes ainsi que des moyens d'apprentissage de chroniques aptes à mettre en œuvre

25

30

une méthode d'apprentissage automatique de chroniques fréquentes d'un journal d'alarmes, de fréquences supérieures ou égales à un seuil de fréquence minimum f_0 réglable et de durée maximum T réglable, et aptes à
5 transmettre les chroniques obtenues, caractérisé en ce qu'il comporte en outre :

- un module de sélection et de regroupement de séquences d'alarmes apte à recevoir un journal d'alarmes et apte à sélectionner et à regrouper des
10 séquences d'alarmes du journal d'alarmes, et apte à former un groupe de séquences d'alarmes similaires et à transmettre ce groupe; et,

- un module de fabrication d'un journal d'alarmes partiel à partir des alarmes d'un groupe de
15 séquences d'alarmes similaires reçu du module de sélection et de regroupement de séquences d'alarmes du journal d'alarmes, le module étant apte à transmettre le journal d'alarmes partiel obtenu aux moyens d'apprentissage de chroniques; et,

20 - un module de fabrication d'un ensemble de chroniques fréquentes du journal d'alarmes, à partir des chroniques transmises par les moyens d'apprentissage de chroniques, le module étant apte à transmettre les chroniques de l'ensemble de chroniques
25 fréquentes.

1 / 4

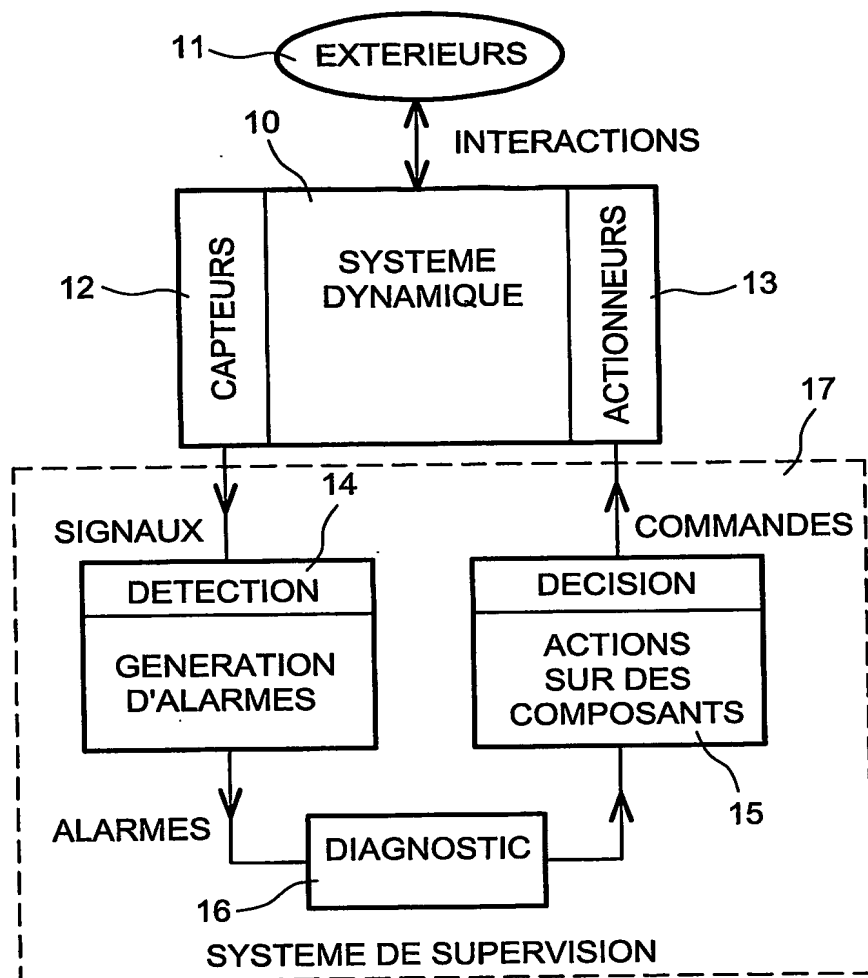


FIG. 1

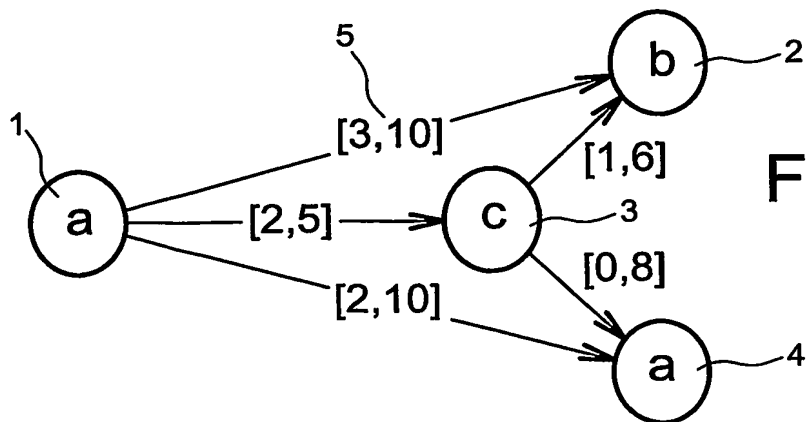


FIG. 2

2 / 4

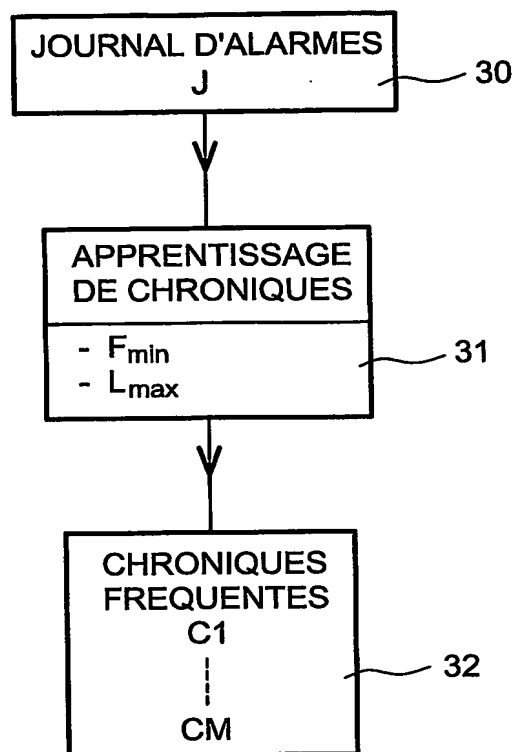


FIG. 3

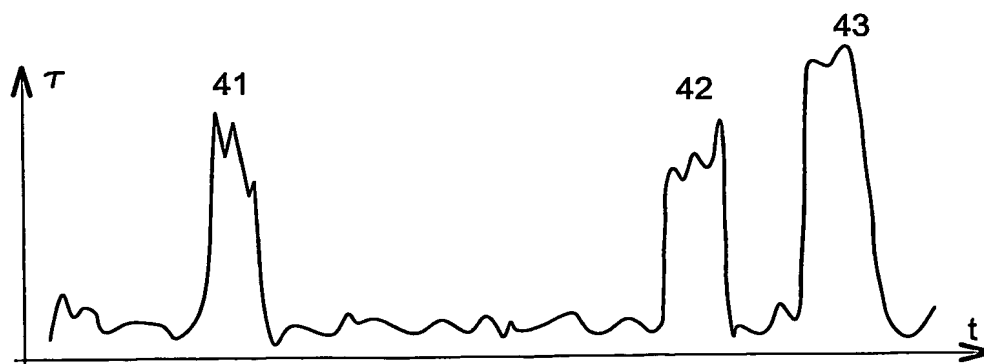


FIG. 4

3 / 4

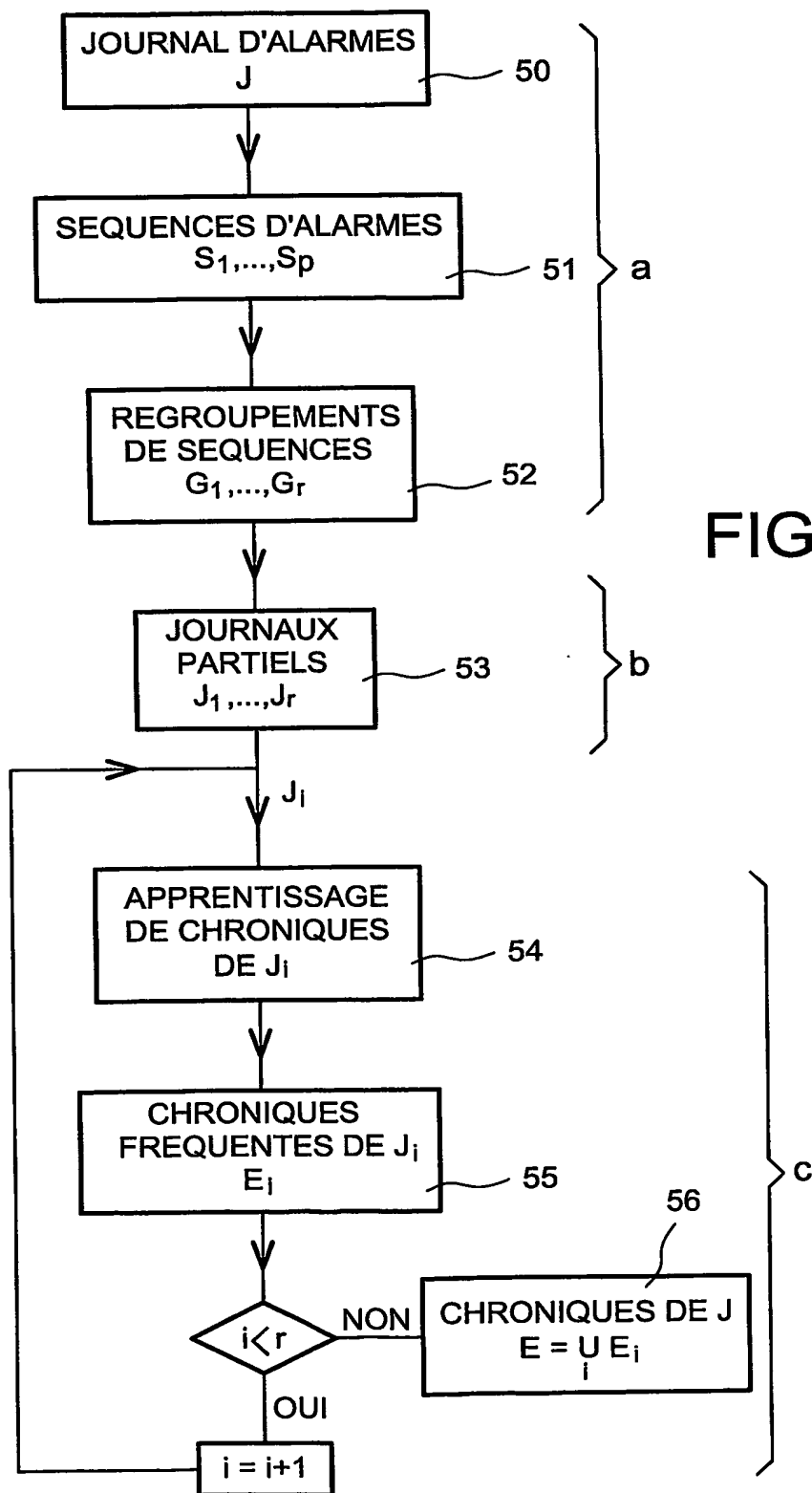


FIG. 5

4 / 4

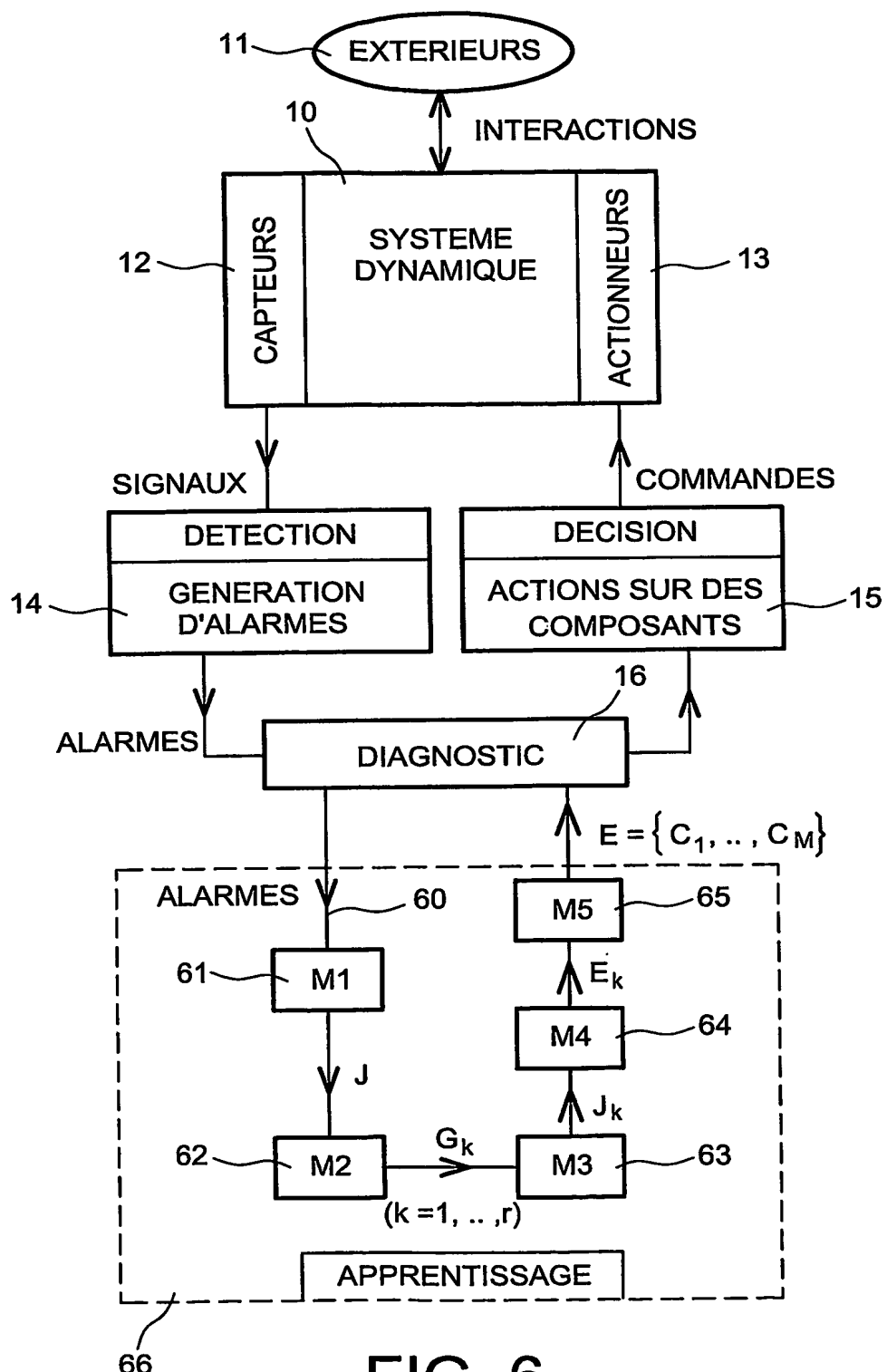


FIG. 6

INTERNATIONAL SEARCH REPORT

International Application No.

PCT/FR2004/050205

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 H04L12/24

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MARILLY E ET AL: "Alarm correlation for complex telecommunication networks using neural networks and signal processing" IEEE, 29 October 2002 (2002-10-29), pages 3-7, XP010611939	1-13,15
A	abstract; figures 1-4 page 4, column 2, line 20 - page 6, column 2, line 24	14
X	LIPPERTS S: "Enabling alarm correlation for a mobile agent based system and network management - a wrapper concept*" IEEE, 28 September 1999 (1999-09-28), pages 125-132, XP010354974	1
A	abstract page 128, column 2, line 17 - page 131, column 1, line 14	2-15
	----- -/-- -----	

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

X document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

Y document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

& document member of the same patent family

Date of the actual completion of the international search

15 October 2004

Date of mailing of the international search report

28/10/2004

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Stergiou, C

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/FR2004/050205

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	ABOELELA E ET AL: "Switching theory approach to alarm correlation in network management" IEEE, 8 November 2000 (2000-11-08), pages 452-461, XP010527472 abstract page 455, column 1, line 20 - page 456, column 1, line 16; figure 4 -----	1-15
A	FR 2 821 508 A (FRANCE TELECOM) 30 August 2002 (2002-08-30) abstract page 10, line 10 - page 14, line 33 -----	1-15

Information on patent family members

Ref/FR2004/050205

Form PCT/ISA/210 (patent family annex) (January 2004)

A. CLASSEMENT DE L'OBJET DE LA DEMANDE
CIB 7 H04L12/24

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE

Documentation minimale consultée (système de classification suivi des symboles de classement)
CIB 7 H04L

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si réalisable, termes de recherche utilisés)
EPO-Internal, WPI Data, PAJ

C. DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie *	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	MARILLY E ET AL: "Alarm correlation for complex telecommunication networks using neural networks and signal processing" IEEE, 29 octobre 2002 (2002-10-29), pages 3-7, XP010611939	1-13,15
A	abrégé; figures 1-4 page 4, colonne 2, ligne 20 - page 6, colonne 2, ligne 24	14
X	LIPPERTS S: "Enabling alarm correlation for a mobile agent based system and network management - a wrapper concept*" IEEE, 28 septembre 1999 (1999-09-28), pages 125-132, XP010354974	1
A	abrégé page 128, colonne 2, ligne 17 - page 131, colonne 1, ligne 14	2-15
	----- -/--	

☒ Voir la suite du cadre C pour la fin de la liste des documents

☒ Les documents de familles de brevets sont indiqués en annexe

* Catégories spéciales de documents cités:

A document définissant l'état général de la technique, non considéré comme particulièrement pertinent

E document antérieur, mais publié à la date de dépôt international ou après cette date

L document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)

O document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens

P document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

T document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention

X document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément

Y document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier

& document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée

15 octobre 2004

Date d'expédition du présent rapport de recherche internationale

28/10/2004

Nom et adresse postale de l'administration chargée de la recherche internationale
Office Européen des Brevets, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Fonctionnaire autorisé

Stergiou, C

C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
A	ABOELELA E ET AL: "Switching theory approach to alarm correlation in network management" IEEE, 8 novembre 2000 (2000-11-08), pages 452-461, XP010527472 abrégé page 455, colonne 1, ligne 20 - page 456, colonne 1, ligne 16; figure 4 -----	1-15
A	FR 2 821 508 A (FRANCE TELECOM) 30 août 2002 (2002-08-30) abrégé page 10, ligne 10 - page 14, ligne 33 -----	1-15

Renseignements relatifs aux membres de familles de brevets

181/FR2004/050205

Formulaire PCT/ISA/210 (annexe familles de brevets) (Janvier 2004)